# A Forecasting Support System dedicated to Temperature-Controlled Goods Hauling

Wilfried Despagne

# A Forecasting Support System Dedicated to Temperature-Controlled Goods Hauling

Wilfried Despagne

*Lab-STICC research laboratory (UMR CNRS 3192)*
*European University of Brittany*
*Centre Yves Coppens, BP 573, F-56017 Vannes, France*
*and*
*STEF-TFE, Cold chain logistics made in Europe*
*Boulevard Malesherbes, F-75008 Paris, France*

**Abstract**

This article describes an operational research problem. A firm specialized in temperature controlled transportation wants to optimize the planning of its human and material resources using short-term activity forecasting. The challenge is to find a unique forecasting model adapted, without any human intervention, to the specific needs of 57 of the company's offices. To do so, the company has been collecting data for five years. Mathematic algorithms for forecasting time series are used to analyse the data. The goal of the work is to combine these tools to extract the maximum amount of determinist information that should be anticipated. The introduction presents the problem and its economic context. It is followed by a description of the chosen process and arguments to defend those choices. The adopted solutions are inventoried. A forcasting support system interface is presented. Finally, the conclusion refers to courses of study.

keywords: time series, forecasting support system, supply chain, carriage

The authors would like to thank Joanna Ropers for her help with the English of this paper.

# 1 Introduction

An assessment of temperature controlled transportation activity forecasting is presented here. It has helped in acquiring a formalized point of view of the problem and an automatization of the procedures.
Temperature controlled transportation is an activity which consists in transporting goods requiring a temperature ranging from $-25°C$ to $+15°C$. The goods are essentially perishable foodstuff, meat products, seafood, fruits and vegetables, diary products, frozen goods, as well as plants and medicine. The point that these goods have in common is that they are subject to the very strict requirements that make up the "cold chain". The "cold chain" is a process that

helps you keep a product at a low temperature. The cold slows down the propagation of micro organisms. The law imposes regulations for it. The May 9, 1955 decree (regulating the hygiene of foodstuffs directly given to consumers) and the July 20, 1998 decree (setting the technical and hygienic conditions as applied to the transportation of foodstuffs) oblige manufacturers to follow these regulations. However, along with the constraints of keeping a set temperature linked to the cold chain, are those dealing with the transportation activity. Here again, lawmakers regulate the amount of time a lorry driver can spend working, as well as regulating the driving authorization. These are just a few examples of the many laws that exist, which can give you an idea of the extensive range of legislative constraints in the transportation domain, not to mention in the refrigerated transportation domain. Without the management of material and human resources, these constraints would cause functioning costs to explode.

The foodstuff domain involves producers, manufacturers, distributors, and transporters, all of which make up the Supply Chain [Ayadi(2005)]. The constraint consists in maintaining the freshness of the products between their place of origin and their place of distribution, thus imposing on the different links of the chain to use just-in-time methods. Use-by dates are only a few days for poultry cuttings and 21 days for yogurts. These dates control the fresh product supply chain and demand a fast pace in distribution. In her dissertation, Charlotte [Terrolle(2004)] states that manufacturers answer the daily orders of supermarkets to supply their warehouses. The quantities of provisions fluctuate daily according to supermarket check-outs. Every evening, the in-store sales data are centralized in the warehouse so that the orders can be sent out the next day. This automatic replenishment makes up 80% of grocery sales and 98% of fresh product sales [Terrolle(2004)]. To maintain this rhythm and fill up shelf space without keeping stock, the actors must forecast their sales. Logically, supermarkets should share their monthly, as well as daily, sales forecasts with food-processing manufacturers who will then send the information on to the transporters. However, this would be against the rules of competition. Supermarkets are scared of giving too much information to manufacturers who could take advantage of this strong hold to raise prices. Therefore, the actors in the supply chain each have their own forecasts.

For this study, we propose a sales forecasting system for a carrier of temperature-controlled goods. Suppose the carrier is a large group that has a network of fifty or so hubs. To define this network we use [Branche(2006)]. A hub has a platform (or quay) on which the goods are received. The quays allow the carrier to prepare orders, as well as sort and label them to deliver them all over France in less than 48 hours. In the jargon of this domain, it is a carrier in *A* for *B*. Generally, there is one hub per region. On the one hand, it is in charge of picking up the goods from regional clients to inject into the network which will deliver it to the right place. These three carriers are called, in order, “pickup”, “forwarding”, and “delivery”. On the other hand, the hub receives the network goods to deliver to the sales locations of its customer catchment area.

As for all the other links of the foodstuff supply chain, the carrier has to work with the just-in-time (JIT) method. In more than 90% of the cases, the hub receives the consignor’s transportation orders in less than 3 hours before the goods are picked up. To summarize the issue, the

transportation orders come in by fax or EDI (Electronic Data Interchange, [Seiersen(2006)] at 7a.m. to go out between 8a.m. and 12p.m. according to the destination. The delivery must be made between 6p.m. and 10p.m. anywhere in France. Due to special offers (sales promotions), the quantities to be transported vary between 1 and 10 from one day to the next. How should such cases be handled when the clients do not give information on these specific periods?
The carrier must, therefore, be able to adapt to the demands of the clients. To assist with this goal, the carrier must set up a forecasting system to help manage resources [Seiersen(2006)]. The system must be able to anticipate the weight of the goods to be transported and number of waybills to be fulfilled. These two pieces of information, foreseen for a 15-day horizon and for daily periodicity, helps to anticipate the material and human numbers on the quay, as well as the number of articulated lorries to make available. For simplification's sake, we will use the terms of activity forecasting, and of weight and waybill number anticipation. Thus, the activity forecasting aims at furnishing crucial elements for:

- managing material and human resources,
- optimising pickup, forwarding, and delivery regulations,
- formalising short, middle, and long-term behaviour of clients,
- obtaining high level of service,
- limiting dependence on uncertainty.

To homogenize procedures, the carrier needs a forecasting system able to adapt to the specificities of its different hubs. Forecasts must be easy to check, user-friendly and accessible through a web interface on the internet [Cluzel(2006)]. Finally, the direction objective is to reach a daily forecast error rate which is lower than 5%.
To answer the traceability requirements imposed by Europe (EC text n° 178/2002), to improve profitability, as well as service to the client, the group uses a Data Warehouse. It is an Oracle© data warehouse that has in stock all of the information relative to parcels transported or stocked by the group. The data come from an information system called GTI (Gestion Intégrée du Transport - integrated transportation management). It includes the computer application of the group and feeds the database, including several giga-octets of information. The reference information is about transportation. It gives access to information on loading and unloading: product, logistic unit, place of loading and unloading, date, hour, third party consignor, third party consignee, type of packing, weight, packing categories ... the database is updated daily with the data from the previous day. All of this information helps to trace the history of transported volumes since 2000. It is what makes up the population on which the forecasting will be based.
This article present a forecasting method which combines endogenous and exogenous methods [Bourbonnais and Usunier(2007)]. The method is aimed at anticipating, daily and on a 15-day horizon, the values of three chronicles, the sum of two of which equals the third one. Two algorithms for decomposition of the chronicle are used. The estimations obtained according

to two algorithms are combined. The top-down aggregation technique is used (see § 2.5) to satisfy the constraints of the three time series. Finally, a presentation of the FSS (Forecasting Support System) interface is given. A part of the following results were announced in the article [Despagne(2008), Despagne(2010)].

# 2 The Model

## 2.1 Overview

Our goal is to model three time series which represent the flow of goods out of quay. The goods are transported to another quay, forwarding, or delivered to their final destinations, delivery. The sum of both gives the total amount of foodstuffs dealt with at that quay. For a precise definition of these three notions see article [Fabbe-Costes(1999)]. The carrier extracts daily the data on both quantities to be forecasted, number of waybills and associated weights. Observation is over 5 years of the data history.
The chosen statistical model is to understand the activity of carrier as much as possible by breaking it down. Suppose a hub's activity is determined by three components: deterministic factors, stochastic factors, and unobserved factors. Deterministic factors (seasonality, bank holidays, sales promotions) are particular in that having been observed in the past, their forthcoming dates are known. Stochastic factors can also be broken down into observed variables (strikes, gain or loss of a client portfolio), but their forthcoming dates are unknown. Unobserved factors are exceptional, new events (compulsory liquidation of main rival), the past, present and future values of which are unknown. The proposed model creates a copy of this history. For each one, it uses an appropriate method to extract the deterministic factors before applying an "autoprojectiv" model to estimate the stochastic variables. The two forecasts obtained after recomposition are linearly combined to give the minimum forecast error. The difference between the model results and observations is due to unobserved factors.
Let the triplet $(X_t, Y_t, Z_t) \in \mathbb{R}^3$ representing the values at date $t$ of the chronicles of "forwarding", "distribution", and "total". The variables are subject to the constraint of $X+Y=Z$. These values are calculated from other time series ($U_{X,t}$, $U_{Y,t}$, $U_{Z,t}$) by $X_t = \omega_1 U_{X,t}$, $Y_t = \omega_2 U_{Y,t}$ et $Z_t = U_{Z,t}$, $\omega_1$ and $Z_t = U_{Z,t}$, $\omega_1$, and $\omega_2$ are determined to satisfy the constraint. The other time series are calculated by:

$$U^1_{E,t} = T_t S^1_t (\beta^1 F)_t V^1_t \varepsilon^1_t \tag{1}$$

$$U^2_{E,t} = T_t S^2_t (\beta^2 F)_t V^2_t \varepsilon^2_t \tag{2}$$

with $E \in \{X, Y, Z\}$, $T$ the tendency, $F$ the binary vector corresponding to the calendar events, $\beta$ their weights, $S$ the seasonal coefficients, $V$ the stationary process and $\varepsilon$ the white noise. To study the components independently from the others, we use the Naperian logarithm function.

$$ln(U^1_{E,t}) = ln(T_t) + ln(S^1_t) + ln((\beta^1 F)_t) + ln(V^1_t) + ln(\varepsilon^1_t) \tag{3}$$

$$ln(U^2_{E,t}) = ln(T_t) + ln(S^2_t) + ln((\beta^2 F)_t) + ln(V^2_t) + ln(\varepsilon^2_t) \tag{4}$$

The results obtained by equations (3) and (4) are combined in equation (5).

$$U_{E,t} = \lambda ln(U^1_{E,t}) + (1-\lambda) ln(U^2_{E,t}) \tag{5}$$

## 2.2 Modelling deterministic elements

### 2.2.1 The tendency

Hub activity has two tendencies, an intra-annual tendency and an inter-annual tendency. The intra-annual tendency is the activity of a hub between January and December. The inter-annual tendency is the long-term one. After substantial end-of-the-year expenses, households start saving again in January. This is why activity is high in December and falls in January. Furthermore, in the economic sector, tendencies are slow and progressive says [Burtschy and Menendian(1980)]. These observations led us to choose a linear tendency by lumps of annual periodicity explains [Vaté(1993)]. The equation is written:

$$T_t = \theta t + \varphi An(t) + cste$$

with $An(t)$ the year corresponding to date $t$ ; $\theta t$ representing the inter-annual tendency, $\varphi An(t)$ the intra-annual tendency.

### 2.2.2 Seasonality

The transported quantities create a superposition of oscillatory movements over weekly and daily time periods, therefore showing double seasonality that is estimated so that it can be erased from the chronicle. The weekly seasonality has 53 coefficients and the daily seasonality has 313 working days. The weekly one is due to periods of fluctuating activity which are caused by external events, such as the weather, school holidays, or holiday seasons. As they are included in the seasonality, it is no longer necessary to analyse them individually. It would be ideal to disregard the number of weeks to keep only the distances in relation to calendar events. Thus, we would no longer have to be preoccupied by the fact that during year $A$, a specific bank holiday fell on week 13 while during the following year it was on week 14. This is an idea to pursue. For the moment, we include 53 weeks, the two ends of which are corrected according to the number of days they have. Week 01 is the one that includes the first Monday of the year. If the 1st of January is not on a Monday, then that day and all the following days until the next Monday are part of week 0. Therefore, week 0 and the last week of the year (52) are the only two that could not include 6 working days. They are weighted according to the number of days they have.

The daily seasonality, which is very pronounced, is due to a distribution of the activity over the 6 working days of the week. This distribution depends on each hub. On Saturdays, for example, activity is reduced to a strict minimum, but also varies according to the supplying days of wholesalers and supermarkets. Indeed, supermarkets make up 80% of the transported volumes. There are numerous methods to make seasonal adjustments. They are advantageous in that

they delineate the activity to those who make the decisions. The seasonal coefficients show the difference between the observed average value for week $i$ and day $j$ in relation to the tendency. For sales forecasting, there is no ideal method. Using the idea that two are better than one, then why not apply two methods to keep one combination of results, as suggested by [Bates and Granger(1969)] and [Schnaars(2006)], according to the minimisation criterion of error variance (see § 2.4). The first seasonal adjustment method is the moving average one. The second is decomposition by linear regression.
The moving average method [Brockwell and Richard(1996)] lets you estimate seasonal coefficients through the following steps,

- calculate the series of central moving averages,
- calculate the difference between the observed values and the moving average,
- normalise the differences to obtain the seasonal coefficients.

This method is applied the first time to correct weekly variation and the second time to correct daily variations. The results obtained are seasonal coefficients $S^1$ of equation (1). The second decomposition method is the one proposed by [Buys-Ballot(1847)]. It consists in finding coefficients $S^2$ of equation (2) by OLS (Ordinary Least Squares).

$$\begin{aligned} ln(U^2_{E,t}) - ln(T_t) = \gamma_1 S_t^{2,1} + \gamma_2 S_t^{2,2} + \gamma_3 S_t^{2,3} + \gamma_4 S_t^{2,4} + \gamma_5 S_t^{2,5} + \gamma_6 S_t^{2,6} + \\ \Phi_0 SS_t^{2,0} + ... + \Phi_{52} SS_t^{2,52} + \zeta_t \end{aligned}$$

The time serie reduced of its tendency, is decomposed into a series of seasonal components corresponding to 6 days of the week, to 53 weeks of the year, and to process $\zeta_t$. Seasonal components $p$ are binary variables, $p$ seasons of the year. The binary variable equals 1 when the datum relates to the considered season and 0 otherwise.

### 2.2.3 Calendar Events

The following events are applied to both series ($ln(U^i_E) - ln(T) - ln(S^i)$, $i = \{1,2\}$), which have been adjusted for season variations and tendency, and which have been obtained by the moving average and Buys-Ballot methods. The loss of one day of activity will create recuperation of that amount of activity throughout the neighbouring days. For example, if Thursday is a holiday, then there can be an increase in activity on Monday in anticipation or Friday due to lateness. Often, a holiday has foreseeable consequences over a 9-day period ($J-4, \ldots, J, \ldots, J+4$). The consequences are different according to the holiday, the day of the week, and the given hub.
Recuperating the loss of a day of work is variable whether the day is a Monday, Tuesday, or another day. If it is a Saturday, there is little activity to make up, whereas a Monday is a busy day. If the holiday is on a Friday, supermarkets anticipate it and request the delivery of twice the quantity the Thursday before. Deliveries boom on Thursday and forwarding increases on

Wednesday for deliveries from $A$ to $B$. If it happens on a Monday, supermarkets will anticipate a little on the previous Friday, but mainly recuperate on the Tuesday. If the holiday is a Thursday, it is highly likely that Friday's activity will be reduced as employees will go on long weekends. The previous Wednesday will be all the busier.

A holiday often coincides with a celebration which causes an increase in consumption by households and consequently an increase in the transportation activity. However, this increase is variable whether it is All Saints' Day or Christmas. Hubs are not all equal when faced with an activity increase due to a holiday. Let us take a hub located near a chocolate factory. The factory supplies all the stores in France with chocolate for Easter. Supplying starts as early as the month prior to the event and monopolises a large part of the hub's resources. Other hubs do not have a similar client and undergo less of an increase than they see at other times of the year. They could transport lily of the valley for Mayday all over France.

Sometimes, the fourth day after a holiday is also the second day prior to another holiday. This is what happens in May between Mayday (May 1st) and VE Day (May 8th). In this case, it is difficult to separate the effects of both holidays.

To try to separate the 4 phenomena caused by a holiday we follow a method inspired by [Longstaff and Wang(2004)]. To do so we need the following information:

- title of the holiday,
- day of the week (Monday, ..., Saturday),
- day of the week of the 4 previous and 4 following days,
- distance of the 8 days around the holiday $(-4,-3,-2,-1,1,2,3,4)$.

This gives a binary table with 20 columns, one to identify the day of the week, eight named between $-4$ and 4 to indicate the distance of the effected day from the holiday and eleven others to identify the holidays. The holidays used are: New Year's Day, Easter, Mayday, VE Day, Ascension Day, Whit Monday, Bastille Day, Assumption Day, All Saints' Day, Armistice Day, and Christmas. T lines represent the number of records in the chronicle. This matrix, called $F$, is composed of binary variables allowing for the adjustment of the regression model on a season adjusted series: $W = \beta F + \xi$, with $W$ as the season adjusted time serie. A comparison between the estimation variance and the error variance (Fischer test) will help to point out the most discriminatory variables.

## 2.3 Stochastic element modelling

The previous procedures have helped extract the tendency, the seasonal components, and the foreseeable events from the time serie. The obtained time series ($V$) has a cause and effect relationship between the observation of a date $t$ and previous observations ($t-1$, $t-2$, $t-3$, $t-4$, $t-5$, $t-6$). To model and foresee this series, we use simple exponential smoothing [Box and Jenkins(1970)] which has the advantage of having automation potential. $\widehat{V}_{T+1} = (1 -$

$\alpha)\sum_{j=0}^{T-1}\alpha^j V_{T-j}$
It takes into account past observations (from $T-1$ to $T-6$) and weighs them with smoothing constant $\alpha$, which is estimated in such a way as to minimise the square of the difference between the time serie and the model estimations.
The predicted values are added to the deterministic components that we previously subtracted. After application of the exponential function, we obtain the estimation of the original time serie.

## 2.4 Forecast combination

Due to the use of two decomposition procedures, we get two forecasts. The forecasting error obtained for both methods does not allow for a choice of a superior one out of the two. The combination set-up is proposed by [Bourbonnais and Usunier(2007)]. Its goal is to minimise forecasting error variance which results in a performance prior to individual forecasts. Not dependent on the specification of one model, the combined forecasts are meant to be more robust. The combined forecast $CF$ is a weighted mean of two individual forecasts $IF^1$ and $IF^2$ ; $CF = \lambda IF^1 + (1-\lambda)IF^2$, $\lambda$ is the weighted coefficient, $0 < \lambda < 1$. Let $CFE$, combined forecasting error, $CFE = \lambda IFE^1 + (1-\lambda)IFE^2$, the forecasting error variance is $V(CFE) = \lambda^2 V(IFE^1) + (1-\lambda)^2 V(IFE^2) + 2(1-\lambda)\lambda COV(IFE^1, IFE^2)$. We want to find $\lambda$ which minimises $V(CFE)$, by cancelling out the first derivative in relation to $\lambda$. If the forecasting errors or correlated, the solution is

$$\lambda = \frac{V(IFE^2) - COV(IFE^1, IFE^2)}{V(IFE^1) + V(IFE^2) - 2COV(IFE^1, IFE^2)}$$

else,

$$\lambda = \frac{V(IFE^2)}{V(IFE^1) + V(IFE^2)}$$

.

## 2.5 Forecasting adjustment and forecasting distance

The time series of weight that has passed through the quay in “forwarding”, “delivery”, and in “total” are forecasted separately. Thus, the total must equal the sum of “forwarding” and “delivery”. To make the forecasts coherent, the numbers must be adjusted. Due to their importance, the “total” numbers are less variant, thus easier to forecast. Using the principle that the forecasting error of the sum is not as high as the accumulated errors of “forwarding” and “delivery”, we have chosen to keep the forecast of the “total goods passed through the quay” series to correct the other two. This technique is called top-down aggregation. The adjustment procedure is the following

- let $\widehat{X}_{t+h}$ be the forwarding weight forecasted for date $t+h$,
- let $\widehat{Y}_{t+h}$ be the delivery weight forecasted for date $t+h$,

- let $\widehat{Z}_{t+h}$ be the total weight forecasted for date $t+h$.

We are looking for $\alpha$ and $\beta$ so that $[\widehat{Z}_{t+h} - (\alpha\widehat{X}_{t+h} + \beta\widehat{Y}_{t+h})]^2$ is minimal, under constraint $1,5 > \alpha > 0,5$ and $1,5 > \beta > 0,5$. The constraints indicate that the "forwarding" or "delivery" forecasts cannot be wrong for more than 50%. The Quasi-Newton method was chosen to solve the optimisation problem. To accompany the forecasting values, we calculate the bilateral forecast distance on a probability level of 95%. The forecasting distance can be useful when a decision is difficult to take. It can, for example, decide on the allocation of another lorry on its rounds. Furthermore, the forecasting distance can give idea of how much trust can be put in a forecast. The greater the distance, the less stable the forecast.

# 3 Application and results

Let there be a transport hub, 60% of the activity of which is in delivery, as compared to the 40% in forwarding. Between January and March, the activity is stable, around 750 tonnes (from Monday to Friday). The holidays in April and May create a great deal of change in the flow rate. The weight of transported goods can go from less than 10 tonnes on a holiday to more than 1110 tonnes the two previous days. Excluding Saturdays and holidays, the activity from April to May comes very close to 1000 tonnes/day. June is a slack period. With an activity of around 900 tonnes/day and some holidays, July and August are busy months. September - October is a period that once again is calm like the beginning of the year with a mean less than 800 tonnes/day. Finally, the activity progresses regularly throughout the month of December to reach a peak of 1114 tonnes five days before Christmas.
Table 1 compares the results obtained for the model with the actual observations. The learning sample is the record of weights prior to date $T$. $T$ varies between 01/01/07 and 07/31/07 with a six-day span. We forecast activity $\widehat{X}_{T+i}$, $\widehat{Y}_{T+i}$, $\widehat{Z}_{T+i}$ from dates $T+1$ to $T+6$ and we compare the forecast to the chronological series $X_{T+i}$, $Y_{T+i}$, $Z_{T+i}$ actually carried out for $i = \{1,..,6\}$.
The evaluation indicators are the relative absolute error (RAE) and the relative dispersion (RD).

| Forwarding | | Delivery | | Total | |
|---|---|---|---|---|---|
| RAE | RD | RAE | RD | RAE | RD |
| 8% | 10% | 9% | 12% | 7% | 9% |

Table 1: Confidence Indicators.

The indicators refer to the mean of observed values. The results of the model are wrong 7% of the time according to the mean of the "total goods passed through the quay" series. The errors are likely to vary more or less 9%.
A way to verify the efficiency of the models is to study the forecasting errors. It must follow a white-noise-process behaviour. [Mélard(1990)] recommends using the test of [Ljung and Box(1978)] to test the autocorrelation of forecasting errors.

| Delays | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Forwarding | 0.003 | 0.0003 | 0.0009 | 0.0016 | 0.0039 | 0.0081 |
| Delivery | 0.896 | 0.742 | 0.5382 | 0.6284 | 0.1983 | 0.0007 |
| Total | 0.453 | 0.4718 | 0.6579 | 0.8027 | 0.4224 | 0.0538 |

Table 2: p-value Ljung&Box Test

For the previous example, the results of the test (table 2)[1], for delays varying from 1 to 6, lead us to believe that the model has captured the dependency of the first 6 delays on the "Total" time serie and the first 5 delays on the "Delivery" time serie. Contrariwise, the forecasting errors for "Forwarding" retain the information allowing for the forecast of their next values from the 6 previous ones. We can therefore not conclude that the independence hypothesis is true for "Forwarding" forecasting errors.
It is notable that the results are at least as good, if not better than those obtained from the carrier's existing models (see [Métivier and Jaffrès(2005)]). This model benefits from the fact that it adapts to the time series of the group's 57 hubs. It is intelligent in that for each calculated forecast, its parameters are estimated again from the history. However, the objective of an average yearly forecasting error that is less than 5%, no matter the hub, was not reached. Although, it seems that according to the hub and the day of the week, a forecast which is 95% correct over the 3 previous weeks is attainable.
One might add that the quality of the forecasts vary according to the volatility of the activity's behaviour. A period of stable activity (January to March) allows one to extrapolate the time serie without inciting large errors. However, turbulent periods (April, May) still pose problems because of the holidays (see § 2.2.3). It seems that the method chosen to measure the effect of holidays on the activity is not the most adequate one. To be effective, the method would need more history. As stated, it estimates a loss or gain coefficient per OLS for each of the 4 days previous to and following a holiday. The cycle that will give you the same day of the week for a given holiday is 28 years. However, a cycle of 11 years, even 7 years, could be acceptable. For more information on the distribution of holidays, see the definition given on the website [Wikipédia(2009)]. Moreover, observations are distant in time and household behaviour changes over time.

# 4 The graphic interface: Tip of the iceberg

To enhance the results obtained by the model, a web application has been developed. This is the tip of the iceberg. Indeed, the hidden side of forecasting support systems (FFS) is the mathematical model, as well as a system of digital data exchange. It allows for the creation of a history of good data, the calculation of forecasts, and the updating of databases.

[1] with a risk of 5%, the null hypothesis is rejected when the p-value is lower than 0.05


Figure 1: Horizons's home page

The graphic interface developed in thin client is meant to reproduce forecasting results in a clear and useful manner. The objective is to make this application a decision-making tool. The solution proposed gives a scientific tool through statistics to anticipate demand. More than a forecasting tool, it also lets you analyse the activity of a hub by navigating through time. The graphic options let you visualise at a glance the atypical data, the lows and highs of the activity. The indicators give information on the correctness of past forecasts (see fig. 2). The planning page offers graphic visualisation and planning elements including a proposition of daily productive hours over a horizon of 5 weeks. The forecasting system's graphic interface provides a complete power and easy-to-interpret solution for anticipating and planning the activity of a carrier hub. The forecasting tool called Horizons should make planning simpler, faster and more reliable.

The interface must be able to post within a delay that is acceptable for the user, information extracted from tables which can have over 6 million records.

The home page (see fig. 1) presents the forecasting tool Horizons. A text recalls the objective of the tool and the information that the user will find there. Hypertext links give help pages and present the project. To go to real and provisional data, the user must first identify him/herself.

After authentication, the page model is divided into 3 parts (see fig. 2). A column on the left presents the navigation menus. A graphic part at the top of the page gives a quick, synthetic vision of all the detailed data found underneath. The part which presents the detailed data is placed to the right of the navigation menu and below the graphic frame.

The graphic is interactive. Once the user approaches a point (intersection between days on abscissa and ordinate data value) using the mouse pointer, a rectangular box pops up to give details


Figure 2: Horizons's weights forcast

relative to this point (date, real datum, forecast, correction, difference in percentage according to the real) (see fig. 2). This box comes up red when the forecasting error is higher than 10%, otherwise it is green. The user has the choice to show between 2, 5, and 10 weeks of data. The 10-week choice gives him/her the long-term tendency of the activity. The 5-week choice shows the forecasting errors of the last two weeks and anticipates the activity of the 3 following weeks, for example. Data over two weeks is useful to communicate on the forecasts. The graph and the detailed data of the 2 weeks displayed fill one printable page. It is best to print the results of the previous week and the forecasts for the next week. Displaying the report for all to see gives transparency on the past forecasting errors, but also motivate the teams by showing them the activity they have done. By knowing the rhythm of the previous week, and by visualising the forecasting report, they can estimate the rhythm they should expect for the current week.
The tables of detailed data provide, for each day of the week, the result of the previous year at the same time period (week, day). This information is very important. It is a referent to validate the forecast. Indeed, the calculation engine is still mechanical, it can get jammed. This is why the user must validate the forecasts. The result in $A-1$ is a referent which allows him/her to do so. This indicator also lets one position themselves according to last year's report. Is the activity low, stable or high?
The second information in the table is the forecast. The third line gives the turnover. Every day, the actual numbers from the previous day are available. It is thus automatically compared to the forecast and contributes to the confidence statistics. The fourth line shows the ajusted and validated value per manager.
Maintaining a goal of transparency, the graphic interface of the Horizons tool gives all of the

statistics to describe the performance of past forecasts (see fig. 3). [Bortolotti(2005)] indicates that "forecasts always have uncertainty and inaccuracy. It is important to evaluate the rough estimate and to communicate it to the users." The web page "confidence indicators" (see fig. 3) gives a diagnosis to evaluate the forecasts of weight and number of positions gone through the quay, in forwarding and delivery. This diagnosis is possible for any time range and for any period between January 1, 2007 and today. Through graphic means (histogram of errors) you can quickly see for which days the model was most effective. For a non-statistician, error variance is more accessible through the histogram than the raw data. Numerical statistics (means, standard deviation, percentage of correct forecasts) confirm the forecast performances.
The performance diagnosis aims at making the statistic indicators easy to understand and interpret. After this consultation, the manger is able to form an in-depth opinion of the confidence that can be given to the proposed forecasts.
The Horizons tool graphic interface proposes a collaborative forecasting window (see fig. 4). It can help to fit and validate each of the forecasts proposed on the horizon of a week. A pattern should appear for each correction. Thus, with time carrier collects crucial information to explain the reasons for activity peaks. The correction patterns are chosen from a list: earnings portfolio, losses portfolio, commercial event, holiday, beginning of the month, end of the month, seasonal

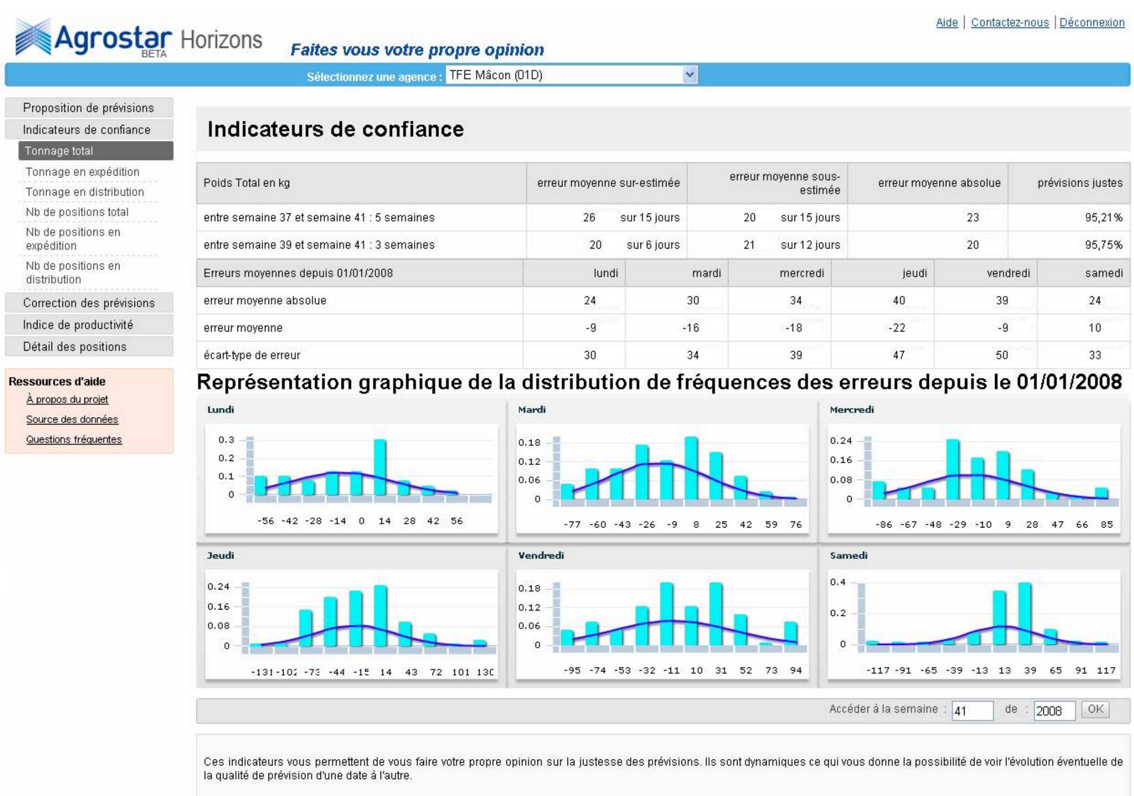


Figure 3: Horizons's confidence indicators

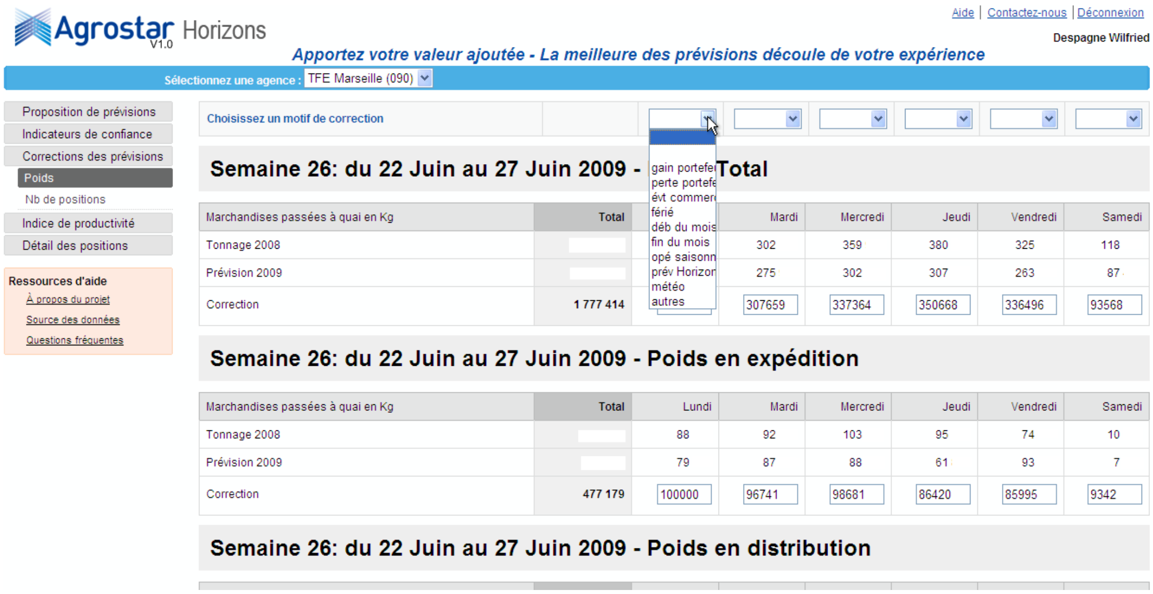


Figure 4: Horizons's collaborative forecasting

operation, Horizons forecast, weather factor, other. This list was discussed with the teams on the field. However, it is not exhaustive and can be added on to. After six months to a year of gathering information, the forecaster can perfect the forecasts. Indeed, knowing that at the same time period the previous year a sales promotion took place, the system should send a warning message to ask if this promotion will happen again, and if it does, takes it into account in the forecasting calculation. As soon as a correction is made, it shows up in the graphic part and all of the users can consult it.

The tool is meant to be a rough draft for planning. This is why the "rate of productivity" section has been added. This section gives, on the same model as a weight forecast, the turnover and the forecast of productive hours and rate of productivity. The rate of productivity is the ratio between the weight and the number of productive hours. It is a productivity indicator for those in charge.

After a first presentation of the interface at pilot hubs, the ergonomy of the site proved reliable, and indeed, is greatly used. However, one question constantly comes up: what is the data source? As long as the users do not understand how the data is calculated and how it is linked to the numerous indicators already in place in the information system, they will not use the forecasting tool. In fact, to make it easier for them to accept the tool, it is crucial that the data posted is reliable, homogeneous, and of controlled quality. Having access to a forecast is not enough, the interpretation and use of the forecast depends on the quality of information given ([Forslund and Jonsson(2007)]).

Two new pages have been created to get the user to accept the presented information. A help page gives the definition of the information, its calculation rule, and its sources. The common denominator of all of the carrier's data is position. This is why another webpage gives, for each hub, all of the positions taken into account in the provided indicators. The information linked to the position is posted. This is a piece of information that the users like very much. From this information, the quay chief knows right away if the position went through his/her quay, at what time, its origin and its destination. From that, the information will be all the easier to accept because he/she can mentally approximate it.

# 5 Conclusion and future research

Forecasting calculations are centralised at the carrier's decision support system service. The algorithm presented in this article updates the 6 chronological series for each of the carrier's 57 hubs every week over a horizon of 28 days. The results are published on the web interface that was specifically created for this purpose. They are presented as indicators or graphs. As the forecasts are on average 90% correct, they are a valuable decision making tool.

The weak point of this model is the excessively large number of parameters compared to the size of the history (5 years). Indeed, for each of the decompositions, there are 2 parameters for the tendency, 313 for the daily seasonality, 53 for the weekly one, 20 for the calendar events and one last one for the exponential smoothing. The advantage is to be able to explain to the decision maker the effect of each of the parameters on the quantity of goods going out of quay. One way to reduce the number of parameters is to estimate the seasonality with a Fourier transform and to classify the calendar events in function to their effects. It is also possible to consider using the correlation that may exist between the time series of the 57 hubs to better the modelling. The correlations come from the fact that hubs make up a network and forward goods between themselves.

A whole other approach would be to do seasonal adjustment through embedded Kalman filters and introduce into the model, in impulsion form, the effects of calendar events. This method is recommended by [Martin(1999)] and applied on forecasts of electricity consumption.

Nevertheless, the economists count more on what they call "collaborative forecasts". Statistics are not enough to get a reliable forecast. With the development of shared management, the results can be enhanced through the validation or the commentaries of different company services (logistics, marketing) and even clients.

# References

[Ayadi(2005)] Ayadi S (2005) Le Supply Chain Management : Vers une optimisation globale des flux. Working paper, Université Catholique de Lyon

[Bates and Granger(1969)] Bates J, Granger C (1969) The combination of forecasts. Operational Research Quarterly 20(4)

[Bortolotti(2005)] Bortolotti R (2005) Analyse du système de prévision des effectifs d'élèves dans l'enseignement secondaire postobligatoire. Note d'information, Service de la recherche en éducation, Genève

[Bourbonnais and Usunier(2007)] Bourbonnais R, Usunier JC (2007) Prévision des ventes, théorie et pratique. Economica

[Box and Jenkins(1970)] Box G, Jenkins G (1970) Time series analysis: Forecasting and control. San Francisco: Holden-Day Inc

[Branche(2006)] Branche F (2006) Transport de messagerie. Revue technique de l'ingénieur dossier n°AG8151 V2

[Brockwell and Richard(1996)] Brockwell JP, Richard AD (1996) Introduction to Time Series and Forecasting. Springer

[Burtschy and Menendian(1980)] Burtschy B, Menendian C (1980) A propos de prévision à court terme de la production industrielle. Revue De Statistique Appliquée 28(2):5–24

[Buys-Ballot(1847)] Buys-Ballot CHD (1847) Les changements périodiques de temperature. Utrecht

[Cluzel(2006)] Cluzel G (2006) Rentabilité d'un système d'information. Approche théorique. Revue technique de l'ingénieur dossier n°AG5310

[Despagne(2008)] Despagne W (2008) Etude préliminaire à un modèle de prévision à court terme de l'activité d'un transporteur sous température dirigée. Modulad 39:95–106

[Despagne(2010)] Despagne W (2010) Construction, analyse et implentation d'un modèle de prévision. Déploiement sous forme d'un système de prévision chez un opérateur européen du transport et de la logistique. Mémoire de fin d'études, Université de Bretagne Sud

[Fabbe-Costes(1999)] Fabbe-Costes N (1999) Système d'information logistique et transport. Revue technique de l'ingénieur dossier n°AG8030 V2

[Forslund and Jonsson(2007)] Forslund H, Jonsson P (2007) The impact of information quality on supply chain performance. International Journal of Operations & Production Management 27:90–107

[Ljung and Box(1978)] Ljung G, Box G (1978) On a measure of lack of fit in time series models. Biometrika 65:297–303

[Longstaff and Wang(2004)] Longstaff FA, Wang WA (2004) Electricity Forward Prices: A High-Frequency Empirical Analysis. The Journal of Finance 59(4):1877–1900

[Martin(1999)] Martin MM (1999) Filtrage de Kalman d'une série saisonnière, Application à la prévision de la consommation d'électricité. Revue De Statistique Appliquée 47(4):69–86

[Mélard(1990)] Mélard G (1990) Méthodes de prévisions à court Terme. Editions Ellipses

[Métivier and Jaffrès(2005)] Métivier C, Jaffrès B (2005) Maintenance de l'outil de prévision Tonnages. Rapport interne, STEF-TFE

[Schnaars(2006)] Schnaars S (2006) An evaluation of rules for selecting an extrapolation model on yearly sales forecasts. Interfaces 16:100–1007

[Seiersen(2006)] Seiersen N (2006) Systèmes d'information logistique. Revue technique de l'ingénieur dossier n°AG5300

[Terrolle(2004)] Terrolle C (2004) Evolution des rapports entre industriels et grande distribution : du partenariat à la satisfaction clients, vers l'émergence de nouvelles stratégies d'achat. Mémoire de fin d'études, Université Paris I Pantéhon-Sorbonne

[Vaté(1993)] Vaté M (1993) Statistique Chronologique et Prévisions. Economica

[Wikipédia(2009)] Wikipédia (2009) Fêtes et jours fériés en France, http://fr.wikipedia.org/wiki/Fance